Coeliac Disease

Edited by
Markku Mäki, Pekka Collin, J.K. Visakorpi

Proceedings of the
Seventh International Symposium on Coeliac Disease
September 5–7, 1996, Tampere, Finland

Coeliac Disease Study Group
Tampere

Gluten and Gliadin: Precipitating Factors in Coeliac Disease

Donald D. Kasarda

U.S. Department of Agriculture, Agricultural Research Service, Western Regional Research Center, Albany, California, USA

The harmful effects of gluten proteins in coeliac disease are possibly the result of an evolutionary collision in which two separately evolving gene systems, through their protein products, began to interact with one another when wheat became an important dietary component of humans (Figure 1). This interaction occurred only about 10,000 years ago when man first began to cultivate wheat and barley, although hunters and gatherers had probably collected wild grains for some considerable time prior to the development of cultivation. The two interacting protein systems brought together by this cultural event were key grain storage proteins (especially the gluten proteins) and certain human proteins involved either in the control of DNA transcription or in cellular signaling functions. Molecular mimicry by gluten proteins or peptides of control proteins may be responsible for disease development through interference with normal signaling and control pathways. Although some of these human proteins have plant counterparts, compartmentalization within the plant presumably prevented undesirable interactions. I shall return to this speculation later.

Evolution of grain storage proteins

The evolution of life began about 4,000 million years ago. Complex organisms had evolved in the seas by about 400 million years ago when the land was first colonized by plants and animals. About 300 million years later, grasses began to evolve within the class of flowering plants, itself a relatively late evolutionary development (1). Eventually, certain species of grasses evolved the proline-rich, glutamine-rich storage proteins (prolamins) that are found in the endosperm of many of our cultivated cereal grains, such as wheat, rye, barley, oats, and corn. The genes encoding these proteins, especially the proline-rich, glutamine-rich domains, apparently developed quite recently in the course of evolution, presumably within the past 100 million years (Figure 1). Fossil evidence for grasses dates back only

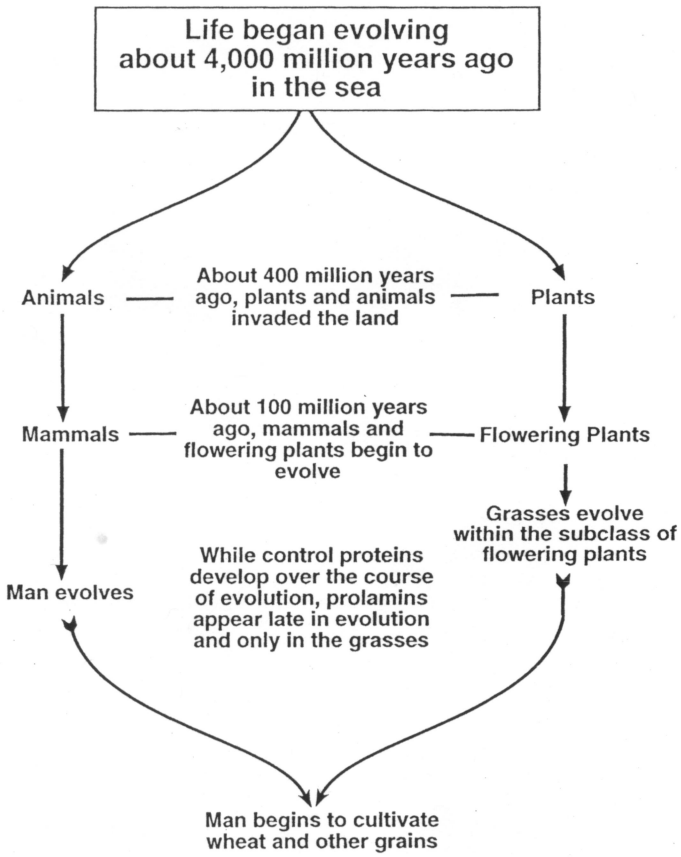

Figure 1. Diagram of hypothetical evolutionary collision between control proteins having proline-rich, glutamine-rich domains and the compartmentalized proline-rich, glutamine-rich prolamins of the grass family through development of agriculture.

to about 65 million years ago (2). The function of grain storage proteins is to provide the developing embryo with a readily utilized source of nitrogen and amino acids upon germination of the seed through enzymatic hydrolysis of storage protein deposits deposited in the grain endosperm during development (3). Presumably, the large amounts of proline and glutamine found in many prolamins serve not only as a good sources of nitrogen, but also enter centrally into the metabolic pathways that lead to other amino acids, nucleic acids, and various other compounds needed for protein synthesis by the new plant (4).

Historically, the term gluten was applied to the storage proteins of wheat grain prepared as a cohesive, elastic mass by washing away the starch granules (Figure 2) from a wheat flour dough (5). Wheat grain is typically 10–15% protein with the remainder consisting mainly of the starch granules. Only the wheat gluten proteins can form a cohesive, elastic dough when flour and water are mixed; this is not readily achievable with pure rye or barley flours. Some semantic confusion has resulted from relatively recent use of the term gluten by physicians and coeliac patients to mean proteins or peptides from any source that are harmful to coeliac patients and by appropriation of the term by the US corn (maize) processors, who refer to certain protein fractions derived from corn as "corn gluten."

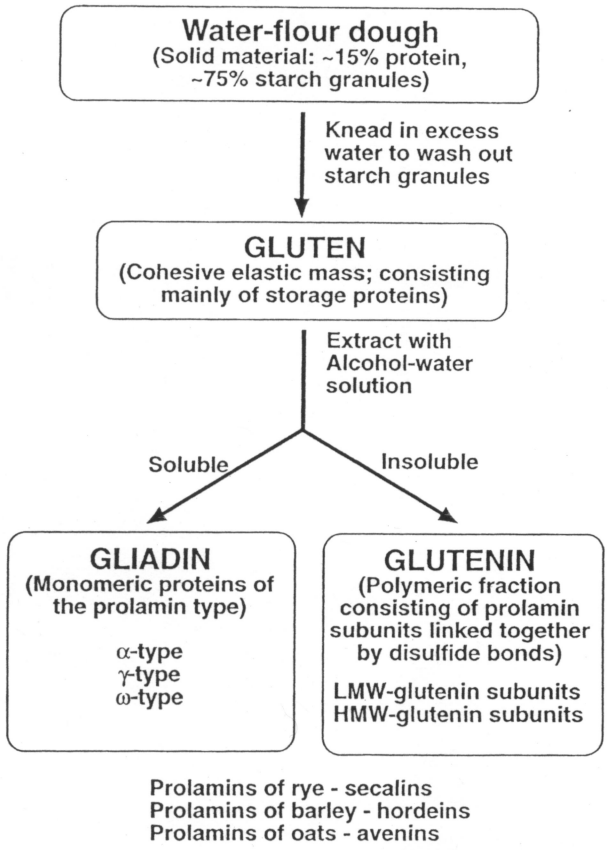

Figure 2. Diagram showing production of wheat gluten from a water-flour dough and fractionation of gluten by solubility into the two roughly equal prolamin fractions, gliadin and glutenin.

Tribal relationships of grass species

At present, it appears that only wheat, rye, and barley (and any species derived from crosses between these species, such as triticale, which combines genetic material from both wheat and rye) have amino acid sequences in their storage proteins capable of initiating processes in those with coeliac disease that may lead to severe intestinal damage. Although the presence of coeliac disease has usually been defined in terms of a flattened mucosa, less severe manifestations are known; an increase in intraepithelial lymphocytes may be the only observable intestinal change in some people (6).

According to plant taxonomy, all the species having harmful proteins are members of the grass family (Gramineae), which is found in the monocots, a major class of the flowering plants, (the other being the dicots). Not all grass species have harmful proteins; rice and corn are considered safe. Grass species more closely related to rice or corn than to wheat, rye, and barley are likely to have grains that are safe for coeliac patients. This is based partly on taxonomy and partly on partial amino acid sequence comparisons (7–9) as there is little controlled scientific testing of grains other than wheat and, recently, oats in

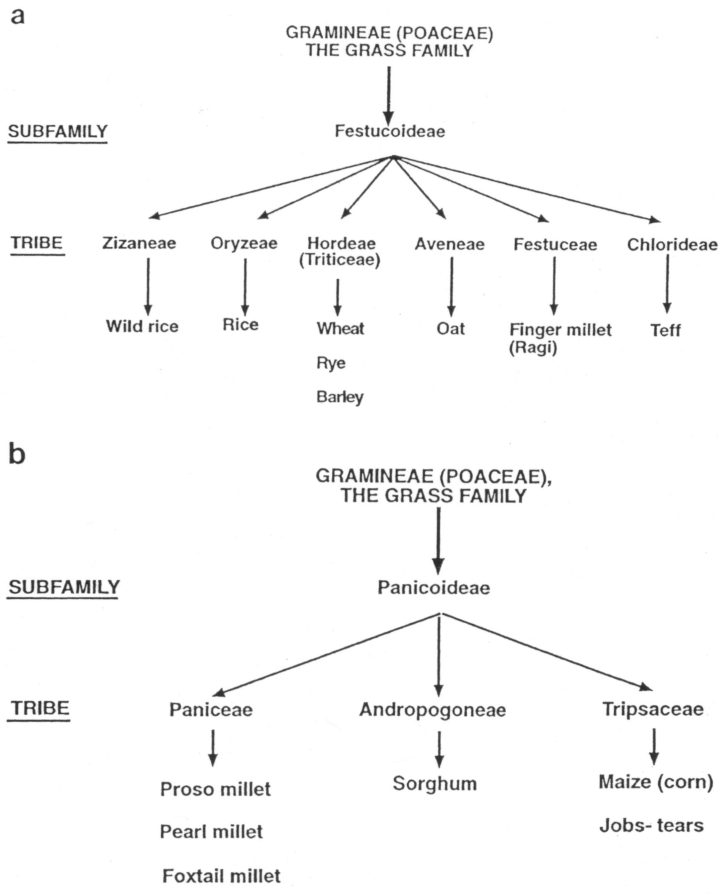

Figure 3. (a) Taxonomic classification by Tribe of harmful species (wheat, rye, barley) and presumably non harmful species (wild rice, rice, oat, finger millet, and teff) of grasses from the Subfamily Festucoideae. Harmful species fall in the Tribe Hordeae (sometimes known as Triticeae). (b) Taxonomic classification by Tribe of presumably non harmful species (various millets, sorghum, maize, and Jobs-tears) of grasses from the Subfamily Panicoideae.

relation to coeliac disease. Figs. 3a and 3b present diagrams showing the relationship at the tribal level of wheat, rye, and barley, the known harmful grains, to some of the grains likely to be harmless to coeliac patients. Wheat, rye, and barley fall in the same tribe, but even oat differs at the tribe level. It is possible that only species found in the tribe Hordeae (sometimes designated Triticeae) have proteins active in coeliac disease. These would include hexaploid wheats, such as normal bread wheats and spelt (or spelta), tetraploid wheats, such as durum wheat used for pasta and *Triticum polonicum* (sometimes called Kamut), and diploid wheats, such as *Triticum monococcum*. Given that only a limited number of grasses have been shown to have proteins with harmful sequences, it seems highly unlikely that species such as buckwheat, amaranth, and quinoa, which are dicots and thereby relatively distantly related to grasses, are likely to have proteins harmful to coeliac patients (8).

All the harmful proteins are proline-rich and glutamine-rich, which provided the basis for the name prolamins given to these proteins by the noted cereal chemist Thomas Burr

The Grain Storage Proteins of Interest Are Notably Proline (P)- and Glutamine (Q)-Rich

Protein type	%P	%Q	%Q + P
α-gliadins	17	38	55
ω-gliadins	30	50	80
LMW-glutenin subunits	15	38	53
avenins (oats)	8	33	41
zeins (maize)	9	21	30

[Percentages (moles/100 moles) are approximate; they vary for specific components of a given type]

Figure 4. Molar percentages of proline and glutamine in various storage prolamins of wheat, oats, and maize.

Osborne (10). Oat prolamins (avenins) and corn prolamins (zeins) are also proline- and glutamine-rich, although having somewhat lower proline content in comparison with wheat, rye, and barley prolamins (Figure 4). The avenins and zeins do not appear to be harmful in coeliac disease (11,12), although the safety of oats has not yet achieved complete acceptance, perhaps because of conflicting earlier results (for discussion see 7,8). These differences in prolamin activity may result from the clear differences in sequence that occur among the proline-rich, glutamine-rich regions of prolamins from different species and for different classes of prolamins within a species.

Gliadin and glutenin fractions

Gluten is frequently divided into two main solubility fractions, the gliadins and the glutenins (Figure 2). The gliadins are monomeric proteins whereas the glutenins, which are responsible for dough elasticity, are made up largely of polymers of proteins linked by disulfide bonds. Any fraction prepared by solubility alone is, however, almost certainly contaminated with proteins of the other fraction, which must be considered in evaluating early work. All gliadin and glutenin proteins have considerable amounts of proline and glutamine and the regions most rich in proline and glutamine are made up of somewhat

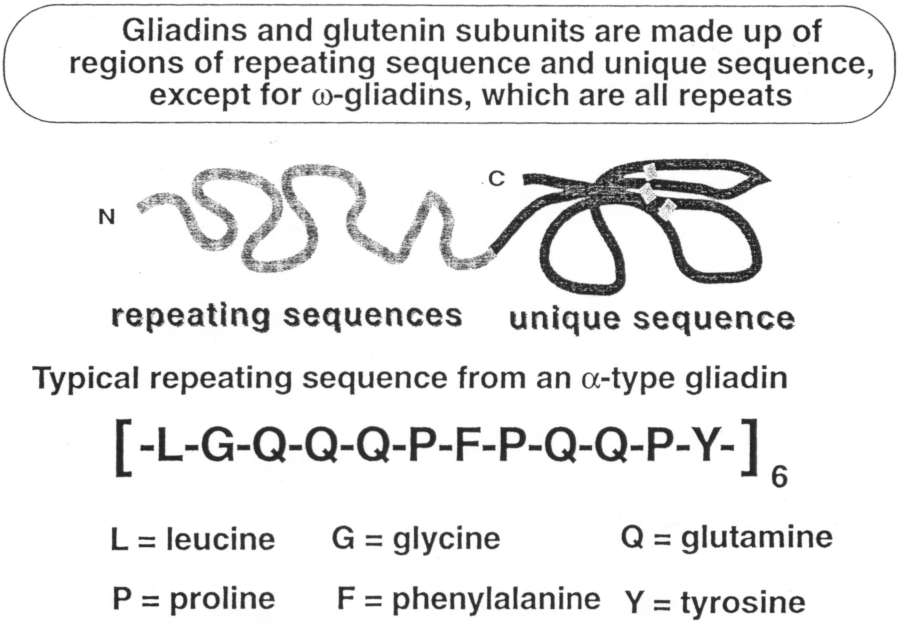

Figure 5. Example of a repeating sequence from an α-type gliadin.

imperfect repeating sequences. An example of a repeat from an α-type gliadin is illustrated in Figure 5. The α-type gliadins also include domains made up of entirely of glutamine residues. Although the size of the polyglutamine repeat domains varies somewhat among the different forms of α gliadin, the larger domain usually incorporates about 18 residues (13). The gluten proteins are coded by multiple genes that show various degrees of divergence depending on the particular type of gene. In the cultivar Cheyenne, there are upwards of 150 genes for the α-type gliadins alone (13), although only about half are expressed—the remainder are pseudogenes.

Early studies were in agreement that activity in coeliac disease resided mainly in the gliadin fraction, and this has been supported by later studies, that indicate activity for all three main structural types (α-, γ-, and ω-) of gliadins (14). These early studies came to mixed conclusions on the activity of the glutenin fraction (for reviews see 7,8). On the basis of sequence similarities to gliadins, the predominating (in glutenin) low-molecular-weight subunits may well be active in coeliac disease. Furthermore, subunits almost identical to gliadins are incorporated into the glutenin fraction by disulfide bonding (15) and consequently there is almost certain to be some toxicity in this fraction.

Gliadin peptides

Various studies have provided *in vitro* evidence for activity in coeliac disease of peptides derived from α- and γ-type gliadins (16–18). In the study of De Ritis et al. (17), we pointed out that two short amino acid sequences, QQQP and PSQQ (where Q represents glutamine, P, proline, and S, serine) were common to the peptides we found active by the organ culture approach, and absent from the two inactive peptides, but indicated that the presence of

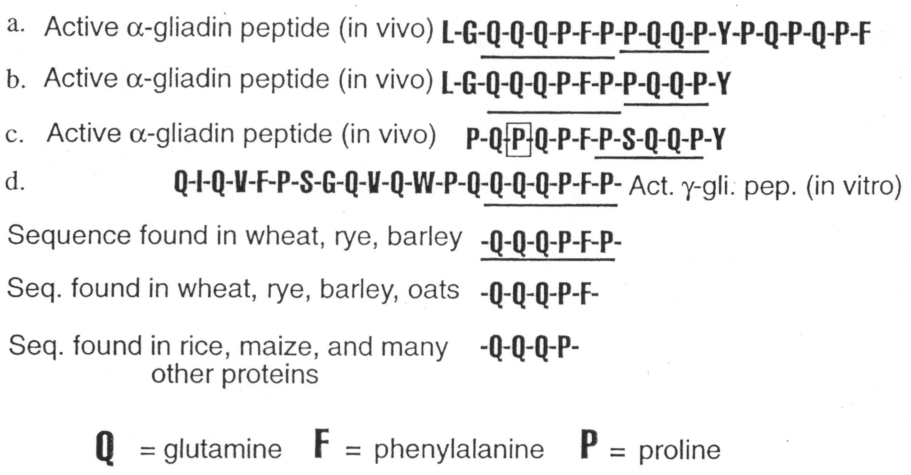

Figure 6. Amino acid sequences of peptides tested for activity in coeliac disease and selected sequences from wheat, rye, barley, and oat prolamins. Sequence a is from Sturgess et al. (19). Sequences b and c are from Marsh et al. (21). Sequence d is from Fluge et al. (18). Potential core sequences are underlined in sequences a, b, c, and d.

these short sequences was not evidence of activity inherent in these short sequences, even when they are parts of larger sequences.

Sturgess et al. (19) provided *in vivo* evidence for activity of a peptide corresponding to amino acid residues 31–49 of an α-type gliadin, for example, A-(α)-gliadin (20), upon instillation of this peptide into the small intestines of several coeliac patients. The amino acid sequence of this peptide is shown in Figure 6(a). Because both *in vitro* and *in vivo* evidence supports the activity of this particular peptide, this sequence is fairly certain to represent an active peptide although it is not likely to be a unique active peptide. Recently, Marsh et al. (21) presented evidence that a truncated version of this peptide, residues 31–43 (LGQQQPFPPQQPY), and a peptide corresponding to residues 44–55 of α-gliadin (PQPQPFPSQQPY) are active as well. These latter sequences are also shown in Figure 6 (b and c).

Relationship of gliadin protein sequences to other proteins

Kagnoff et al. (22) compared the sequence of α-(A)-gliadin with viral protein (and other) sequences in a protein database maintained at the University of California, San Diego. They found a significant similarity between part of the A-gliadin sequence and the E1b protein associated with infection by adenovirus type 12 (Ad 12). Implicit in this work was the possibility that molecular mimicry of the viral protein by gliadin proteins might activate the immune system, triggering coeliac disease in susceptible individuals. This work indicated a possible basis for an environmental contribution to development of coeliac disease, presumably acting in concert with genetic susceptibility. Although there is evidence for and against the hypothesis that Ad 12 is involved in the onset of coeliac disease

(see ref. 23), the work of Kagnoff et al. (22) pointed toward the possible importance of molecular mimicry in the process, regardless of whether or not Ad 12 is involved.

Subsequently, Lähdeaho (24), on the basis of a search of the European Molecular Biology Laboratory (Heidelberg) databases, reported additional sequence similarities in several other proteins, including human annexin and the 100 kD late protein associated with adenovirus type 5. Two other notable sequence similarities have been reported: BM-180 (25), a basement membrane protein found in extracts of lacrimal and parotid glands, had an N-terminal sequence identical to that of α-gliadin through 13 residues (a possible connection to circulating anti-gliadin antibodies in Sjögrens syndrome was suggested). A rather weaker similarity was reported for calreticulin and the α-gliadin sequence by Karska et al. (26), who suggested calreticulin as a possible autoantigen in coeliac disease.

Results of database searching

To complement and extend previous work, I searched the databases available through the National Center for Biotechnology Information (Bethesda, Maryland) by using the BLAST search algorithm of Altschul et al. (27) to look for both identities and similarities between gliadins and other proteins based on either protein or nucleic acid sequences. The goal was to provide further clues to molecular mimicry as the basis for gliadin activity, taking advantage of more recent information about the activity of specific sequences, such as A-gli 31–49 and A-gli 31–43. This work was motivated in part by an intent to provide clues to other potential disease mechanisms in the event of the failure (see refs. 28,29) of the currently favored mechanism involving MHC class II binding of gliadin peptides and presentation of these peptides to T-cell receptors.

An enormous amount of information was retrieved from searches involving only selected gliadin sequences, mainly the N-terminal half of the α-(A)-gliadin sequence (20), which includes residues 31–49, and the N-terminal half of a γ-gliadin sequence. Consequently, I shall be able to present here only a few representative results and provide some general impressions.

I found that the sequences QQQP and PSQQ, which were common to all the active gliadin peptides in the study of De Ritis et al. (17), are found in a very large number of different proteins. Furthermore, the QQQP sequence is included in specific storage proteins of maize (30) and rice (31), and is thereby unlikely to be sufficient in itself to produce the activity of the A-gli 31–49 peptide (which does not include the PSQQ sequence). It seemed notable, however, that these sequences are especially common in proteins involved in cell differentiation, development, and transformation, DNA transcriptional activation, intracellular communication, or neuronal signaling. The reason for this may arise from the common occurrence of proline-rich, glutamine-rich, or proline- and glutamine-rich domains in control and receptor proteins (for example, see refs. 32,33), where, in at least some cases, these sequences are known to bind to other proteins, or to other parts of the same protein.

Extending the QQQP sequence by the next amino acid in the A-gli 31–49 sequence to QQQPF (where F stands for phenylalanine) excluded most proteins other than those of wheat, rye, barley, and oats. Furthermore, no maize or rice proteins were included in the output for this sequence. Next, extending the test sequence by yet one more amino acid,

making it QQQPFP, differentiated oat proteins, which at least on the basis of current sequence information did not have this sequence, from the wheat, rye, and barley proteins, in which this sequence could be found. The relevant sequences are summarized in Figure 6.

Thus, if oat avenins are not toxic, as seems likely from recent work (11,12), this difference may indicate that an active core sequence may include the QQQPFP motif (although oat proteins do not have an additional suspect sequence motif, PQQP, or possibly PQQPY, *vide infra*) and the core sequence might actually involve a combination of these two sequences. The QQQPFP sequence was present in α-type and γ-type gliadins, in low-molecular-weight glutenin subunits, in rye secalins, and in barley hordeins according to the database search. As far as I know, this sequence has not been tested for activity either *in vitro* by the organ culture technique or *in vivo* by feeding or instillation. When I use the term core sequence, I mean to imply that it may be a necessary part of an active peptide. Other residues, however, specific, or non-specific, may be necessary to fill out the size for proper binding to a receptor site on another protein or to a protein complex. In contrast to my emphasis here on the QQQPF sequence, Marsh et al. (21; and personal communication) favor either PQQP or PSQQP as the most likely motif or core sequence involved in coeliac disease. Confident analysis must await further results. Synthetic variations on the theme of the α-gliadin peptide comprising residues 31–43 would certainly be valuable in clearly establishing the key residues in this known toxic sequence.

Possible relationship of the QQQPFP sequence to signaling proteins

There were only four proteins, other than the wheat, rye, and barley prolamins indicated above, that included an identical QQQPFP sequence in their primary structures. These included three hypothetical proteins (obtained as part of genome sequencing efforts) from the yeast *Saccharomyces cerevisiae* (Accession No. gi 849162), from the nematode *Caenorhabditis elegans* (Accession No. gi 466502), a protein from a cyanobacterium *Synechocystis* sp. strain PCC6803 (Accession No. gi 1001474) , and a Golgi apparatus membrane sialoglycoprotein from the rat called MG-160 (Accession No. gi 498341). The MG-160 protein showed homologies (or similarities) to a fibroblast growth factor receptor and to E-selectin (selectins are cell adhesion molecules that play a role in the inflammatory response). The *C. elegans* protein showed similarity to N-acetylgalactosaminyltransferase. No homologies were noted for the cyanobacterium protein. The *S. cerevisiae* protein seemed of special interest in that it was indicated to have homologies to several Src-homology 3 (SH3) containing proteins, including human growth factor receptor-bound protein 2 (Grb-2), which is a cytoplasmic signalling molecule (adaptor protein).

The SH3 domain was recognized in the Src family protein tyrosine kinases (33). These kinases were originally recognized as transforming factors associated with Rous sarcoma virus, but have cellular counterparts. They are involved in many cellular processes, including T-cell receptor signaling. The SH3 domains, which are not limited to the tyrosine kinases, are involved in intramolecular regulation processes and tend to interact with regions of proteins or peptides that form left-handed polyproline II type helices. The finding that the *S. cerevisiae* protein itself contains some proline-glutamine rich sequences

may indicate that this connection may be merely a chance variation with no special significance. However, the complex interactions of molecules such as Grb-2 with other proteins are not completely understood and it is at least possible that the proline-glutamine rich sequences of Grb-2 themselves interact with SH3 domains of other proteins or even those of Grb-2 itself. Interaction of a proline-rich sequence with a neighboring SH3 domain of the same polypeptide chain has been reported for a tyrosine kinase of the Tec family (34).

Various proline-rich motifs that bind to SH3 domains have been described, but P-x-x-P, where x is any residue, appears to be an important consensus sequence involved in the binding process (35,36). Although proline is not essential for the formation of polyproline II type structure, a higher proline content would tend to stabilize this conformation (35). Peptides bound to MHC class II molecules have also been shown to possess or acquire a polyproline type II helical structure (36–38). Furthermore, the P-x-x-P motif is a common one in gluten proteins, especially in the form of -P-Q-Q-P-. This latter combination of amino acids can also be found in Grb-2 (Accession No. gi 1421089), although this sequence is found at the border of one of the SH3 domains and has not been implicated so far in the functioning of the adaptor protein.

Control and signaling proteins that have domains rich in proline, glutamine, or both amino acids developed in plants and animals prior to evolution of the proline-rich, glutamine-rich grain storage proteins. The grain storage proteins are likely to be isolated from interaction in plants with proline- or glutamine-rich signaling/control proteins by compartmentalization in protein bodies during their synthesis in the grain endosperm (3). Ingestion of gluten proteins by man upon development of plant cultivation about 10,000 years ago, when wheat became an important part of the diet in certain geographical areas, may have broken down this separation between gluten and signaling/control proteins. If proline-rich, glutamine-rich peptides from gluten proteins generated by the digestive process interact with or pass through the epithelial membrane (that this occurs seems generally accepted), these peptides might mimic intrinsic signaling molecules by binding to receptor sites at various locations throughout the body, thereby disrupting normal signaling cascades. Such molecular mimicry might be involved in the development of coeliac disease and other gluten sensitive diseases, such as dermatitis herpetiformis, recurrent oral ulceration, nephropathy, and dementia (39,40). This speculation does not, however, explain the strong correlation with the MHC class II locus or why most people are not affected by the proposed molecular mimicry.

The proline-rich nature of at least parts of gluten proteins may contribute to this proposed molecular mimicry by providing a strong tendency for peptides containing proline-rich sequences to form the polyproline type II helical conformation. This tendency may promote binding to receptor grooves, for example, those characteristic of SH3 domains. In Figure 7, I show a computer molecular model of the coeliac-active α-gliadin peptide corresponding to residues 31–43 in the polyproline type II conformation. This conformation imparts a corkscrew-like structure to the main polypeptide chain that has a three-residue repeat. Every fourth residue appears on the same side of the polypeptide chain in the left-handed polyproline type II helical conformation. For example, note the three proline side chains marked by arrows. These correspond to prolines 36, 39, and 42, which are part of both the QQQPFP motif and the PQQP motif. These prolines might bind in hydrophobic pockets defined by aromatic residues and proline residues of an SH3 domain receptor

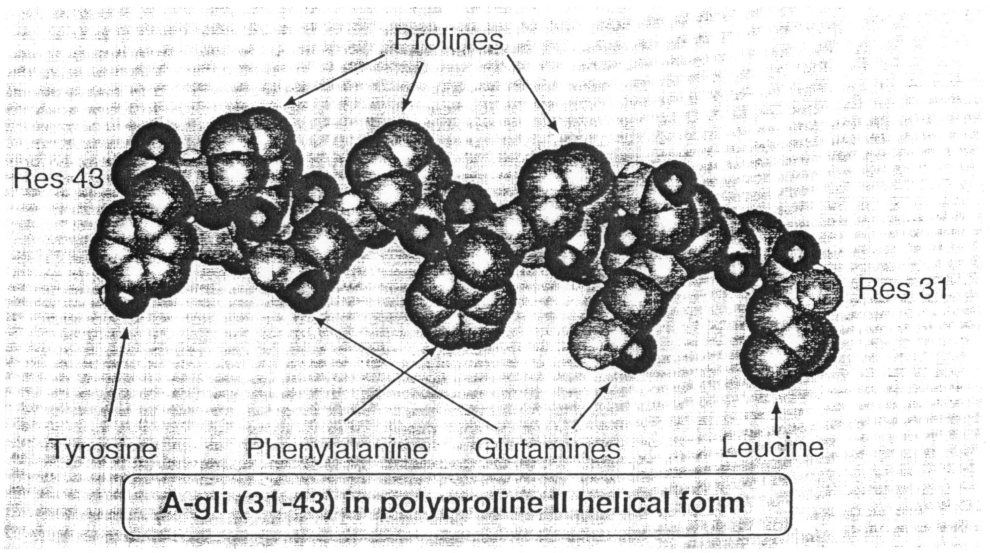

Figure 7. Computer molecular model of A-(α)-gliadin peptide comprising residues 31–43 shown in the left-handed polyproline II helical conformation. Molecule constructed and energy minimized with the Quanta-CHARMm software from Molecular Simulations, Inc. (San Diego, CA).

groove (35). The gaps in between the prolines of the gliadin peptide may fit side chains of receptor groove polypeptides much as the teeth of interlocking gears fit together (35). It is conceivable that the combination of QQQPFP and PQQP is required for gliadin peptide activity in coeliac disease and that the conformational structure might define the activity of an active peptide as much as the actual sequence.

If a polyproline II structure turned out to be important for binding of the active gliadin peptides to a receptor domain, only certain residues whose side chains are directly involved in interactions with the receptor side chains might necessarily be invariant whereas others might vary without affecting activity. The active peptide of Marsh et al. (21) shown in Figure 6 (c) does not have an exact QQQPFP motif or an exact PQQP motif; the apparently deviant amino acids in this sequence might not affect the formation of polyproline II helical structure; they might not be involved in the interaction of this peptide with some as yet unknown receptor site.

There is no direct evidence for interaction of gliadin peptides with signaling molecules, such as those containing SH3 domains. Because there is no clear picture as yet about how gliadins initiate harmful effects in coeliac disease, I simply put these speculations, and the sequence similarities on which they are based, forward as worthy of further consideration in the search for the mechanisms involved.

Similarities between wheat gliadin sequences and those of other proteins

Searching of the sequence databases for similarities to the N-terminal regions of α-gliadins and γ-gliadins yielded a large number of similarities, although these were weak. A few examples are shown in Figures 8–12.

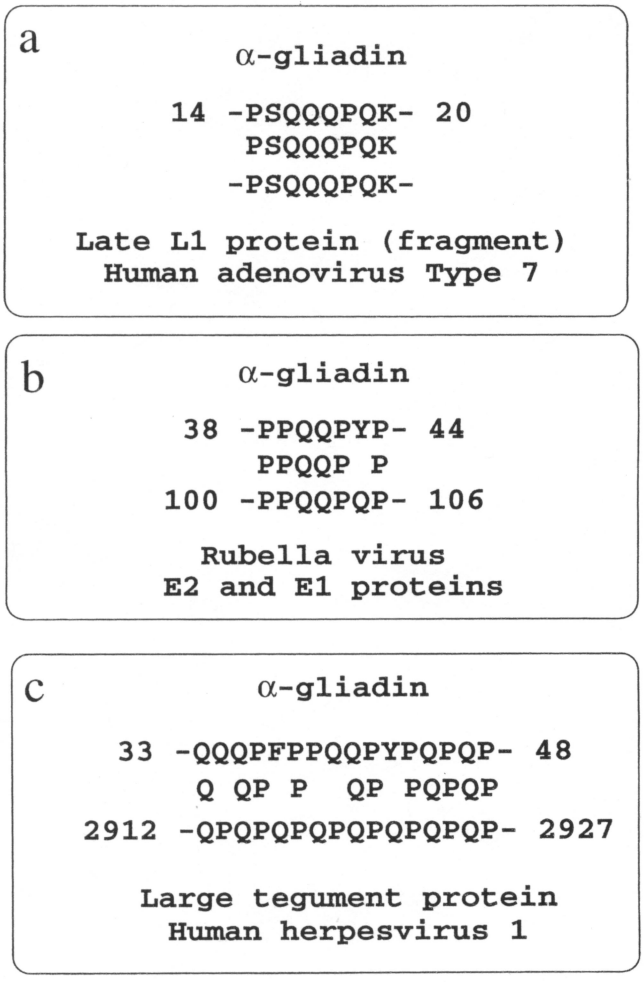

Figure 8. (a) Comparison of the sequence of a fragment from the late L1 protein from human adenovirus type 7 (Accession gi 58533) with residues 14–20 of α-gliadin (sequence from ref. 39); (b) Comparison of residues 100–106 of the E2 and E1 proteins of rubella virus (Accession gi 333973) with A-(α)-gliadin (20) residues 38–44; and (c) comparison of residues 2912–2927 of the large tegument protein of human herpesvirus 1 (Accession gi 135576) with A-(α)-gliadin residues 33–48.

In Figure 8a, an α-type gliadin sequence (41) is compared with a sequence from a late L1 protein of human adenovirus type 7. With eight contiguous identities, these two sequences show similarity equivalent to or greater than was found for the E1b sequence of adenovirus type 12 described to Kagnoff et al. (22). The latter sequence comparison spanned 12 residues with eight identities and some breaks. In Figure 8b, residues 38–44 of A-(α)-gliadin (20) are compared with residues 100-106 of two early proteins associated with rubella virus. There are six out of seven identities and the common sequences include the PQQP motif. In Figure 8c, residues 33–48 of A-(α)-gliadin are compared with residues 2912–2927 of the large tegument protein of human herpesvirus 1. There are 11 out of 16 identical residues in this last comparison. An environmental trigger of coeliac disease is a distinct possibility (22) and viral infection might be the environmental factor involved.

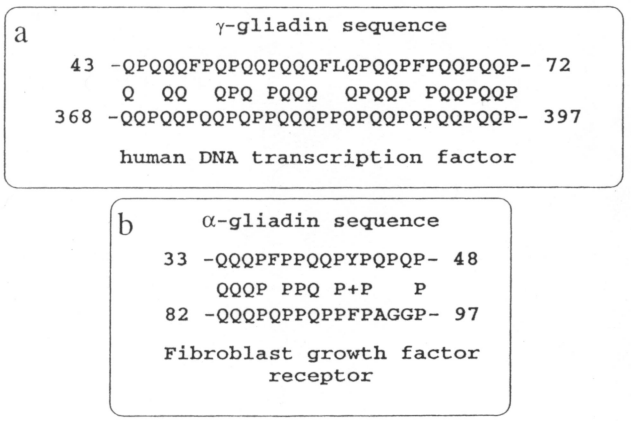

Figure 9. (a) Comparison of residues 368–397 of a human DNA transcription factor (Accession gi 401769) with residues 43–72 of a γ-gliadin sequence (Accession gi 121099; ref. 42); (b) Comparison of residues 33–48 of an A-(α) -gliadin sequence with residues 82–97 of a fibroblast growth factor receptor (Accession gi 1373019).

Accordingly, similarities to viral proteins of viruses that infect humans are worthy of note.

In Figure 9a, residues 43–72 of a γ-gliadin sequence (42) are compared with residues 368–397 of a human DNA transcription factor. Interference with transcription control could come into play in many ways in the development of coeliac disease, perhaps in the activation of T-cells, as just one example. In Figure 9b, residues 33–48 of the A-(α)-gliadin sequence are compared with residues 82–97 of a fibroblast growth factor receptor. Such receptors presumably play a role in the regeneration of tissues, including those of the intestinal mucosa. Gliadin peptides might mimic the receptor, binding to proteins or peptides that normally would interact with the receptor, thereby interrupting a signaling cascade. Fibroblast-derived proteins have been shown to react strongly with circulating IgA from patients with coeliac disease by Marttinen and Mäki (43).

In Figure 10a, residues 96–119, which include the first polyglutamine domain (18 residues) of A-(α)-gliadin are compared with residues 38–61 of interleukin 2 from the mouse. Interleukin 2 is a cytokine secreted by T lymphocytes and stimulates proliferation of activated T lymphocytes. In Figure 10b, the 18-residue polyglutamine domain of A-(α)-gliadin (residues 96-119) is compared with residues 18–35 of the polyglutamine sequence of the human Huntington disease gene. In Huntington's disease, a polyglutamine sequence of about 23 sequential glutamine residues that is characteristic of the normal gene product

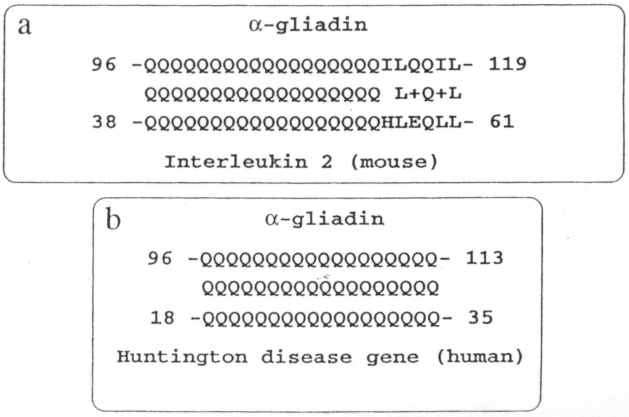

Figure 10. (a) Comparison of residues 38–61 of interleukin 2 from the mouse (Accession gi 349518) with residues 96–119 of A-(α)-gliadin; (b) Comparison of residues 96–113 of the human Huntington disease gene-based sequence (Accession gi 1163054) with residues 96–113 of A-(α) -gliadin.

Figure 11. (a) Comparison of residues 276–292 of a circumsporozoite protein from Plasmodium yoelli (Accession gi 160171) with residues 38–54 of A-(α)-gliadin; (b) Comparison of residues 408–421 of a human serum response factor-related protein 9 (Accession gi 88601) with residues 33–46 of A-(α) -gliadin.

is expanded in the disease-associated protein to about 44 sequential glutamine residues (44,45). In addition to Huntington's Disease, polyglutamine expansion is associated with other neurodegenerative disorders (46).

In Figure 11a, residues 38–54 of A-(α)-gliadin are compared with residues 276–292 of the circumsporozoite protein of *Plasmodium yoelli*, one of the protozoan parasites capable of causing malaria in animals. Although I do not mean to imply a connection between malaria and coeliac disease, parasite infection, like viral infection, is a possible environmental trigger. Antigenic parasite proteins having gliadin-like sequences might trigger coeliac disease in susceptible individuals through molecular mimicry. The presence of proline-rich, glutamine-rich domains in the proteins of several different parasites, none of which, however, is found in humans who live in temperate climes, was noted in a search of the data bases and other parasites may have such domains. For example, *Theileria parva*, which is responsible for East Coast fever in African cattle, has a polymorphic immunodominant protein that includes the tandemly repeated sequence QPEP (47). It is at least possible that some human parasite might also have surface proteins with proline-rich, glutamine-rich sequence that mimic gliadin sequences in such a way as to initiate coeliac disease. Further analysis of parasite protein sequences will inevitably be carried out. In Figure 11b, residues 33–46 of A-(α)-gliadin are compared with residues 408–421 of the human serum response factor-related protein 9. Serum response factors are DNA-binding proteins that play a regulatory role in the transcription of specific genes.

In Figure 12a, residues 33–42 of A-(α)-gliadin are compared with residues 18–27 of avenins from oats that have been shown to react with antibodies found in coeliac serum (48). Although the similarity (in this case, possibly homology) is moderately good between the gliadin sequence and the avenin sequence, the absence of both the QQQPFP motif and the PQQP motif may be significant to the apparent absence of activity in coeliac disease by the oat avenins (11,12). An *in vitro* test of this avenin sequence (perhaps extended

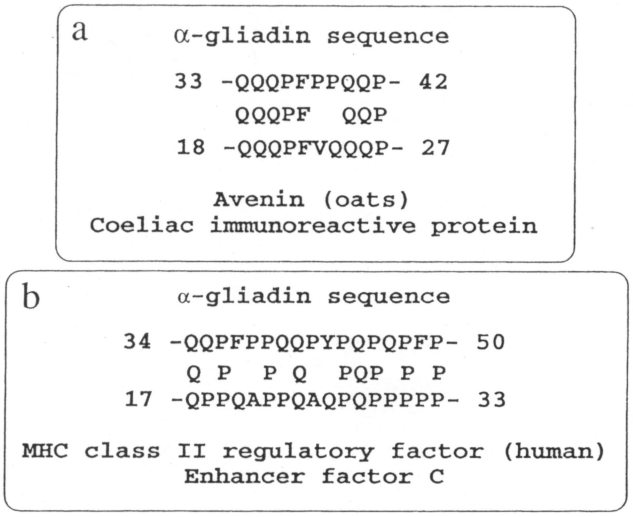

Figure 12. (a) Comparison of residues 18–27 of a coeliac immunoreactive protein (avenin) from oat (Accession gi 322822) with residues 33–42 of A-(α)-gliadin; (b) Comparison of residues 17–34 of a human MHC class II regulatory factor (enhancer factor C) (Accession gi 132518) with residues 34–50 of A-(α)-gliadin.

with a few additional residues to make sure the size is appropriate; a possible candidate being EPFVQQQPFVQQQQ) by the methods of Marsh et al. (21) might go a long way toward eliminating any remaining concerns about the toxicity of oat proteins. In Figure 12b, residues 34–50 of A-(α)-gliadin are compared with residues 17–33 of a human MHC class II regulatory factor (enhancer factor C). The similarity is somewhat weak, but is included as an indication that gliadin peptides might have effects on MHC class II activities that do not involve direct binding to the peptide receptor site of MHC proteins.

Conclusion

No one sequence similarity stands out as being an exceptional candidate for explaining the involvement of gliadin peptides in coeliac disease or other gluten-sensitive diseases, but these similarities indicate a potential for molecular mimicry as a contributing factor, possibly through binding to molecules involved in signaling and control cascades. The databases, already large, are expanding at an enormous rate. As experimental work provides clues to the cellular processes involved in triggering coeliac disease and, in many cases, leading to intestinal damage, searching of the sequence databases for similarities to gliadins may then lead to proteins directly involved in these processes.

Summary

Traditionally, gluten is the term applied to the cohesive, elastic mass of grain storage proteins remaining after starch granules are washed from a wheat flour dough. In this sense, gluten is found only in wheat, but, in coeliac disease, the term has come to indicate grain proteins or peptides that include amino acid sequences harmful to people with coeliac disease. Such sequences are found in barley and rye in addition to wheat. These grains belong to the grass family, although not all grasses have proteins with coeliac-active

sequences—rice, for example. Recent results indicate that oats do not seem to be harmful to coeliac patients. The oat avenins have a truncated core amino acid sequence relative to one found in wheat, rye, and barley storage proteins, and if the lack of toxicity of oat proteins is substantiated, these sequence differences may provide important clues to the nature of the harmful sequence or sequences. Gluten is made up of two major solubility fractions, gliadin and glutenin. Gliadin is definitely harmful. Glutenin is likely to be harmful. A synthetic peptide corresponding to a 12 amino acid residue sequence found in a wheat gliadin has been shown to be active in coeliac disease. Gluten proteins are unusually rich in two amino acids, glutamine (Gln) and proline (Pro). Although the exact mechanism by which gluten proteins or peptides initiate the processes leading to changes in the intestinal epithelium is not known, this compositional peculiarity is almost certainly a factor in coeliac activity. Many proteins that have controlling functions in DNA transcription, cell growth, neuronal signaling, and so forth, have Pro-rich or Gln-rich domains, or both Pro- and Gln-rich domains. Molecular mimicry by gluten proteins and peptides absorbed from the small intestine might result in binding to receptor sites, thereby interfering with normal control pathways. A tendency for gluten peptides to adopt a conformation similar to that of the left-handed polyproline II type helix may be involved in such binding—for example, to the SH3-like domains of tyrosine protein kinases and adaptor proteins or to the peptide receptors of MHC class II molecules.

References

1. Cronquist A. The evolution of flowering plants. Houghton Mifflin, Boston. 1968.
2. Thomasson JR. Paleoagrostology: a historical review. Iowa State Journal of Research 1980;54:301–17.
3. Galili G, Altschuler Y, Levanony H. Assembly and transport of seed storage proteins. Trends in Cell Biology 1993;3:437–42.
4. Miflin BJ, Lea PJ. Amino acid metabolism. Ann Rev Plant Physiol 1977;28:299–329.
5. Beccari J B De Frumento. De Bononiensi Scientarium et Artium. Instituto atque Academia Commentarii 2 (Part 1). Bologna: L Vulpe. 1745;122–7.
6. Marsh MN. Gluten, major histocompatibility complex, and the small intestine. Gastroenterology 1992;102:330–54.
7. Shewry PR, Tatham AS, Kasarda DD. Cereal proteins and coeliac disease. In: MN Marsh, ed. Coeliac Disease. London: Blackwell Scientific, 1992;305–48.
8. Kasarda DD. Toxic cereal grains in coeliac disease. In: C Feighery and C O'Farrelly, eds. Gastrointestinal Immunology and Gluten Sensitive Disease. Proc. 6th International Symp. on Coeliac Disease. Dublin: Oak Tree Press, 1994;203–20.
9. Tatham AS, Fido RJ, Moore CM, et al. Characterization of the major prolamins of tef (*Eragrostis tef*) and finger millet (*Eleusine coracana*). J Cereal Sci 1996;24:65–71.
10. Osborne TJ. The Proteins of the Wheat Kernel. Carnegie Institution of Washington Publication No. 84, Washington, DC. 1907.
11. Janatuinen EK, Pikkarainen PH, Kemppainen TA, et al. A comparison of diets with and without oats in adults with celiac disease. New England J Med 1996;333:1033–7.
12. Srinivasan U, Leonard N, Jones E, et al. Absence of oats toxicity in adult coeliac disease. Brit Med J. 1996;313:1300–1.
13. Anderson OD, Greene FC. The α-gliadin gene family: II. DNA and protein sequence variation, subfamily structure, and origins of pseudogenes. Theoretical and Applied Genetics. 1997;95: 59–65.

14. Ciclitira PJ, Evans DJ, Fagg NLK, Lennox ES, Dowling RH. Clinical testing of gliadin fractions in coeliac disease. Clinical Science 1984;66:357–64.
15. Lew EJ-L, Kuzmicky DD, Kasarda DD. Characterization of LMW-glutenin subunits by RP-HPLC, SDS-PAGE, and N-terminal amino acid sequencing. Cereal Chemistry 1992;69:508–15.
16. Wieser H, Belitz H-D, Idar D, Ashkenazi A. Coeliac activity of the gliadin peptides CT-1 and CT-2. Zeitschrift für Lebensmittel-Untersuchung und-Forschung 1986;182:115–7.
17. De Ritis G, Auricchio S, Jones HW, Lew EJ-L, Bernardin JE, Kasarda DD. *In vitro* (organ culture) studies of the toxicity of specific A-gliadin peptides in celiac disease. Gastroenterology 1988;94:41–9.
18. Fluge Ø, Sletten K, Fluge G, Aksnes L, Elsayed S. In vitro toxicity of purified gluten peptides tested by organ culture. J Pediatr Gastroenterol Nutr 1994;18:186–92.
19. Sturgess R, Day P, Ellis HJ, et al. Wheat peptide challenge in coeliac disease. Lancet 1994;343:758–61.
20. Kasarda, DD, Okita TW, Bernardin JE, et al. Nucleic acid (cDNA) and amino acid sequences of α-type gliadins from wheat (*Triticum aestivum*). Proc Natl Acad Sci, USA 1984;81:4712–6.
21. Marsh MN, Morgan S, Ensari A, et al. In vivo activity of peptides 31–43, 44–55, 56–68 of α-gliadin in gluten sensitive enteropathy (GSE). Gastroenterology 1995;108:A871.
22. Kagnoff MF, Austin RF, Hubert JJ, Bernardin JE, Kasarda DD. Possible role for a human adenovirus in the pathogenesis of celiac disease. Journal Exp Med 1984;160:1544–7.
23. Lähdeaho M-L, Lehtinen M, Rissa H-R, Hyöty H, Reunala T, Mäki M. Antipeptide antibodies to adenovirus E1b protein indicate enhanced risk of celiac disease and dermatitis herpetiformis. Int Arch Allergy Immunol 1993;101:272–6.
24. Lähdeaho M.-L. Homology between Adenovirus E1b protein and A-gliadin. Academic Dissertation, Acta Universitatis Tamperensis, ser A, vol. 430, University of Tampere, Tampere, Finland 1994; also, Lähdeaho M-L, Vainio E, Lehtinen M, et al. Activation of celiac disease immune system by specific α-gliadin peptides, Cereal Chem. 1995;72:475–9.
25. Laurie GW, Ciclitira PJ, Ellis HJ, Pogany G. Immunological and partial sequence identity of mouse BM180 with wheat α-gliadin. Biochem Biophys Res Commun 1995;217:10–5.
26. Karska K, Tuckova L, Steiner L, Tlaskalova-Hogenova H, Michalak M. Calreticulin—the potential autoantigen in coeliac disease. Biochem Biophys Res Commun 1995;209:597–605.
27. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of Molecular Biology 1990;215:403–10.
28. Thurley CA, Kwan WC, Freeman HJ, Gilks CB. T cell receptor gene expression and genotypes in celiac disease. Pathobiology 1994;62:311–8.
29. Johansen BH, Gjertsen HA, Vartdal F, et al. Binding of peptides from the N-terminal region of alpha-gliadin to the celiac disease-associated HLA-DQ2 molecule assessed in biochemical and T cell assays. Clin Immunol Immunopath 1996;79:288–93.
30. Wang S-Z, Esen A. Primary structure of a proline-rich zein and its cDNA. Plant Physiology 1986;81:70–4.
31. Shorrosh BS, Wen L, Sen KC, et al. A novel cereal storage protein: molecular genetics of the 19kDa globulin of rice. Plant Molecular Biology 1992;18:151–4.
32. Gerber H-P, Seipel K, Georgiev O, et al. Transcriptional activation modulated by homopolymeric glutamine and proline stretches. Science 1994;263:808–11.
33. Superti-Furga G. 1995. Regulation of the Src protein tyrosine kinase. FEBS Lett. 369:62–6.
34. Andreotti AH, Bunnell SC, Feng S, Berg LJ, Schreiber SL. Regulatory intramolecular association of a tyrosine kinase of the Tec family. Nature 1997;385:93–7.
35. Lim WA, Richards FM. Critical residues in an SH3 domain from Sem-5 suggest a mechanism for proline-rich peptide recognition. Structural Biology 1994;1:221–4.
36. Alexandropoulos K, Cheng G, Baltimore D. Proline-rich sequences that bind to Src homology 3 domains with individual specificities. Proc Natl Acad Sci USA 1995;92:3110–4.
37. Fremont DH, Hendrickson WA, Marrack P, Kappler J. Structures of an MHC class II molecule with covalently bound single peptides. Science 1996;272:1001–4.
38. Jardetzky TS, Brown JH, Gorga JC, et al. Crystallographic analysis of endogenous peptides associated with HLA-DR1 suggests a common, polyproline II-like conformation for bound peptides. Proc Natl Acad Sci USA 1996;93:734–8.

39. O'Farrelly C. Wheat protein sensitive disease may be independent of intestinal structural changes. In: C Feighery and C O'Farrelly, eds. Gastrointestinal Immunology and Gluten Sensitive Disease. Proc 6th International Symp on Coeliac Disease. Dublin: Oak Tree Press, 1994;169–80.
40. Collin P, Pirttilä T, Nurmikko T, Somer H, Erilä H, Keyriläinen O. Celiac disease, brain atrophy, and dementia. Neurology 1991;41:372–5.
41. Okita TW, Cheesbrough V, Reeves CD. Evolution and heterogeneity of the α/β-type and γ-type gliadin DNA sequences. J Biol Chem 1985;260:8203–13.
42. Scheets K, Rafalski JA, Hegcoth C, Söll D. Heptapeptide repeat structure of a wheat γ-gliadin. Plant Sci Lett 1985;37:221–5.
43. Marttinen A, Mäki M. Purification of fibroblast-derived celiac disease autoantigen molecules. Ped Res 1993;34:420–3.
44. The Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 1993;72: 971–83.
45. Li X-J, Li S-H, Sharp AH, et al. A huntingtin-associated protein enriched in brain with implications for pathology. Nature 1995;378:398–402.
46. Trottler Y, Lutz Y, Stevanin GI, et al. Polyglutamine expansion as a pathological epitope in Huntington's disease and four dominant cerebellar ataxias. Nature 1995;378:403–6.
47. Toye P, Nyanjui J, Goddeeris B, Musoke AJ. Identification and neutralization and diagnostic epitopes on PIM, the polymorphic immunodominant molecule of Theileria parva. Infection and Immunity 1996;64:1832–8.
48. Rocher A, Colilla F, Ortiz ML, Mendez E. Identification of the three major coeliac immunoreactive proteins and one α-amylase inhibitor from oat endosperm. FEBS Lett. 1992;310: 37–40.

Correspondence to: Donald D. Kasarda, U.S. Department of Agriculture, Agricultural Research Service, Western Regional Research Center, 800 Buchanan Street, Albany, CA 94710-1105, USA

Institute of Medical Technology, University of Tampere
P.O. Box 607 (Lenkkeilijänkatu 6)
FIN-33101 TAMPERE, Finland
Phone +358 3 215 7724 Fax +358 3 215 7746
web-site http://www.uta.fi/~llmama/cdstudy/index.htm

ISBN 951-44-4293-8

Printed by Vammalan Kirjapaino Oy, Vammala 1997

Cover figure: The coeliac disease iceberg and spectrum of gluten sensitivity.
By permission of The Lancet; Mäki M, Collin P. Coeliac disease. Lancet 1997;349:1755–1759 (No. 9067).

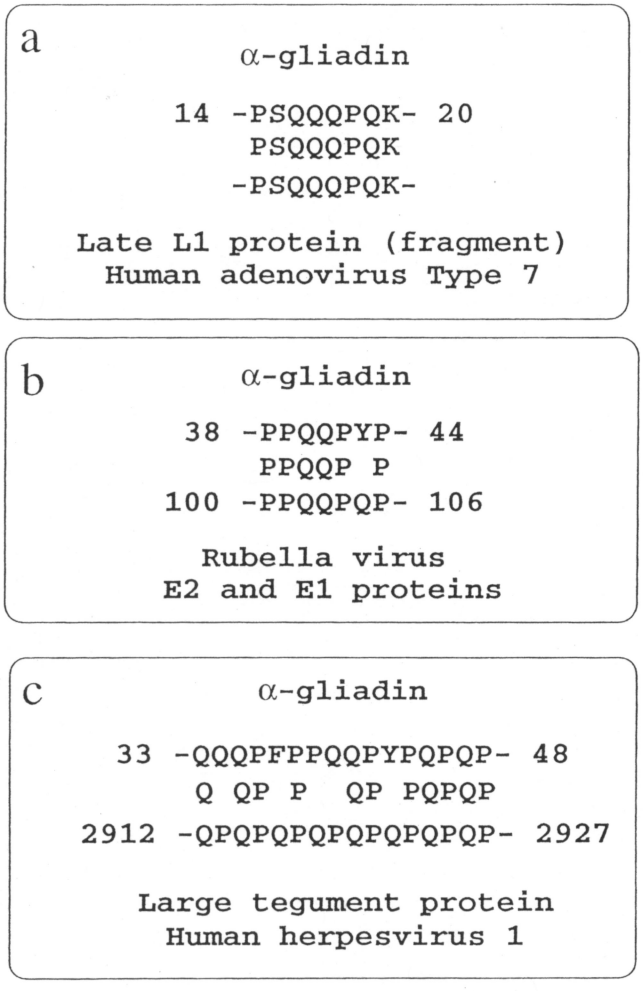

Figure 8. (a) Comparison of the sequence of a fragment from the late L1 protein from human adenovirus type 7 (Accession gi 58533) with residues 14–20 of α-gliadin (sequence from ref. 39); (b) Comparison of residues 100–106 of the E2 and E1 proteins of rubella virus (Accession gi 333973) with A-(α)-gliadin (20) residues 38–44; and (c) comparison of residues 2912–2927 of the large tegument protein of human herpesvirus 1 (Accession gi 135576) with A-(α)-gliadin residues 33–48.

In Figure 8a, an α-type gliadin sequence (41) is compared with a sequence from a late L1 protein of human adenovirus type 7. With eight contiguous identities, these two sequences show similarity equivalent to or greater than was found for the E1b sequence of adenovirus type 12 described to Kagnoff et al. (22). The latter sequence comparison spanned 12 residues with eight identities and some breaks. In Figure 8b, residues 38–44 of A-(α)-gliadin (20) are compared with residues 100-106 of two early proteins associated with rubella virus. There are six out of seven identities and the common sequences include the PQQP motif. In Figure 8c, residues 33–48 of A-(α)-gliadin are compared with residues 2912–2927 of the large tegument protein of human herpesvirus 1. There are 11 out of 16 identical residues in this last comparison. An environmental trigger of coeliac disease is a distinct possibility (22) and viral infection might be the environmental factor involved.